Econometrica, Vol. 59, No. 3 (May, 1991), 859–867

NOTES AND COMMENTS

SOCIAL STABILITY AND EQUILIBRIUM

BY ITZHAK GILBOA AND AKIHIKO MATSUI[1]

1. INTRODUCTION

IN THE FIELD OF NONCOOPERATIVE GAME THEORY, Nash equilibrium (Nash (1951)) has played a central role as a solution concept. In bold strokes, one may discern two major interpretations of Nash equilibrium in the context of rational players.

The first, which is close to the "eductive" interpretation of Binmore (1987, 1988) and the "complete information" interpretation of Kaneko (1987), assumes that the game is played exactly once (if it is a repeated game, the repetition occurs once), and the players have sufficient knowledge and ability to analyze the game in a rational manner. Sometimes it is assumed that all players have consistent hierarchies of beliefs, where the game and their priors are common knowledge. Bayesian interpretation such as proposed by Aumann (1987) advanced this idea to the level that the players have a common prior. From this point of view, however, Nash equilibrium seems far from being satisfactory as it does not satisfy some requirements of "strategic stability."[2] Thus, many studies have been made to refine the concept; among them are Selten (1975), Myerson (1978), Kalai and Samet (1984), and Kohlberg and Mertens (1986). Some studies (see, e.g., Brandenburger and Dekel (1987)) loosen the requirement of common knowledge, but still require some *a priori* knowledge.

The second interpretation is sometimes referred to as the "evolutive" (Binmore) or "naive" interpretation (Kaneko). It does not require that participants in the game know its structure or other facts at the outset. According to this interpretation, a similar situation is repeated many times, and people use trials and errors in choosing better strategies on the basis of information they gradually acquire. A Nash equilibrium is considered as a stationary point in this repeated situation.

At this point, it is worth noting that the price theory of an earlier age such as Walrasian economics shares the basic view of the world with the naive interpretation. It assumes rational participants in the economy but does not assume any common knowledge among participants. They do not know and do not have to know the entire structure of the economy; rather, they observe aggregated signals such as prices on the basis of which they determine their behavior.

This "naive" price theory has solved many economic problems under some appropriate assumptions on the market structure. For example, in a perfect competition model, the assumption of price-takers results in that the participants have (usually unique) dominant strategies as a function of the price signal. The purpose of this paper is to apply similar analysis to general n-person normal-form games. In our model, we assume a large population out of which individual players are randomly matched to play a one-shot normal-form game; hence each one of them may consider oneself a "price-taker" and ignore one's effect on others' behavior.

In price theory and game theory alike, there is interest in the stability of an equilibrium, and more generally, in the dynamics of processes which may or may not lead to an equilibrium. However, in our interpretation of a game, this question seems even

[1] The authors wish to thank the participants of a seminar at Northwestern University as well as Professors Ehud Kalai, Andreu Mas-Colell, Kiminori Matsuyama, Robert Weber, two anonymous referees, and especially Dov Monderer for many useful comments and references. The first author gratefully acknowledges partial support from NSF Grant No. IRI-8814672.

[2] See the discussion in Kohlberg and Mertens (1986).

		Type 2	
		L	R
Type 1	L	1, 1	0, 0
	R	0, 0	1, 1

FIGURE 1

more relevant and unavoidable than in price theory since Nash equilibrium in mixed strategies typically involves nonunique best responses. To support Nash equilibrium in our interpretation we have to assume that a certain portion of the population chooses each specific strategy, while all the population is indifferent among several of them. In other words, even if all players are perfectly rational and the population is at equilibrium, there is no compelling reason to believe it would stay there. There are equally or more probable scenarios according to which every individual plays optimally and yet the behavior pattern moves away from the equilibrium point.

In defining a solution concept on the basis of the naive interpretation, we require it to satisfy the following four qualifications. First, as in a perfectly competitive market, it is assumed that each player is sufficiently small and anonymous, and then may maximize his/her expected utility without getting involved in complicated strategic considerations such as retaliation. Second, unlike a deviation made by a single player, a change in behavior pattern is made in a continuous way. This expresses the intuitive idea that within a small time interval only a correspondingly small proportion of the individuals realize the current behavior pattern and change their strategies. Thirdly, individuals are myopic and choose best response strategies to the *current* behavior pattern. The important consequence of this assumption is that the behavior pattern may form a cycle. Finally, there is a certain limitation in recognizing the current situation. No matter how much information one gathers, it is hard to tell the exact behavior pattern of the society at a given moment.

Similarly to the case of complete information, the concept of Nash equilibrium is not satisfactory as a solution concept when we take the above features into consideration.[3] For example, in the game of coordination, which is shown in Figure 1, there are three Nash equilibria, namely, $([L],[L]), ([R],[R])$, and $(\frac{1}{2}[L]+\frac{1}{2}[R], \frac{1}{2}[L]+\frac{1}{2}[R])$. In the "real world," if the behavior pattern fluctuates toward, say, $([L],[L])$ from the third equilibrium, and if that tendency is observed, then people are likely to follow that behavior. Therefore, the mixed strategy equilibrium of this example is unlikely to sustain itself as a stationary point of some dynamic process. We will propose a new solution concept, called "cyclically stable set," to capture these intuitive ideas, and extend them beyond the mere classification of Nash equilibria to stable and unstable ones.

We first consider the following notion of accessibility, the precise definition of which will be given in the following section: given $\varepsilon > 0$, a strategy profile g is ε-accessible from f if there is a continuous path starting with f and ending in g, such that the direction at each point of the path is a best response to some strategy in the ε-neighborhood of that point; a strategy profile g is accessible from f if there exists a g' sufficiently close to g and ε sufficiently close to zero such that g' is ε-accessible from f. A cyclically stable set (CSS) is a set of strategy profiles such that no strategy profile outside the set is accessible from any strategy profile inside the set, and all the strategy profiles in the set are accessible from each other. In particular, if the cyclically stable set is a singleton, we call

[3] The refined concepts in the context of strategic stability can also be viewed as refinements on the basis of the "naive" or "evolutive" interpretation, in which case they have similar defects to those of Nash equilibrium.

its element a socially stable strategy. We will prove that cyclically stable sets always exist and that each one of them is closed and connected. This new, set-valued solution concept is quite different from various refinements of Nash equilibrium, such as trembling hand perfect (Selten (1975)), persistent (Kalai and Samet (1984)), proper (Myerson (1979)), evolutionary stable strategy (Maynard Smith and Price (1973)), and stable set (Kohlberg and Mertens (1986)). It also differs from the notion of fictitious play (e.g., Shapley (1964)). For a more detailed discussion, see Gilboa and Matsui (1989).

One important feature of CSS's is the independence of sequential elimination of strictly dominated strategies. The situation we have in mind is that all the individuals are so "small" that they do not have to consider the effect of their choices on the distribution of the population, and that all the individuals make no mistakes except that they cannot recognize the present situation precisely (even in that case, their choices are made in a rational manner on the basis of their observation.) In this situation no one should care about strictly dominated strategies, which cannot be chosen at a stationary state. On the other hand, weakly dominated strategy may be present in the support of strategy profiles in a CSS since an individual does not care or does not even know the payoff difference that appears only when other types of individuals take strategies which are not used. Note that Selten's concept of trembling hand perfectness and Kalai and Samet's persistent equilibrium are affected by strictly dominated strategies (see examples in Gilboa and Matsui (1989)).

Another important property is independence of redundant strategies. Note that Myerson's proper equilibrium does not satisfy this property (see an example in Gilboa and Matsui (1989)).

Our model and solution concept are general enough to deal with various random-matching processes. Consider for simplicity a game in which two people are matched. Then the following two cases are distinguished. In the first, the two people matched are from different groups of individuals, say, male and female. In the second, they belong to the same type. In n-person games, in which there are exactly n participants, this distinction does not bother us since each person is assumed to have his/her own identity; on the other hand, in n-type games, which typically involve many participants of each type, information is gathered about types, while the decision makers are individuals. Hence, should two individuals of the same type be matched, each may choose a strategy independently of the other, but the aggregate strategy profile has to be symmetric.

We allow the model to cope with both situations. Our results are stated and proved in a general framework in which a "game" involves the encounter of several (possibly one) individuals of each type. The cyclically stable sets will, of course, depend on the assumptions regarding the identity of types of different players (see Gilboa and Matsui (1989)).

The rest of this paper is organized as follows. Section 2 presents some definitions and notations. Section 3 defines the new solution concept which captures the idea of social stability. In Section 4 we prove its existence and discuss some of its properties. In particular, it is shown that any socially stable strategy, which is defined as the element of a CSS if the latter is a singleton, is a Nash equilibrium and that any strict Nash equilibrium is a socially stable strategy. We also give an example of a game which has no intersection between the set of Nash equilibria and the union of all cyclically stable sets.

2. DEFINITIONS AND NOTATIONS

In a society, equivalently—a "game," there are several types of individuals. Some people are matched randomly to take some actions. In each matching situation, the number of participants from each type is fixed and may exceed one. Therefore, depending on the setting, two individuals of the same type may be matched.

Formally, a game G is described by a quadruple:

$$G = \left\langle I, M, (S_i)_{i \in I}, (\pi_j)_{j \in I} \right\rangle$$

where $I = \{1, 2, \ldots, n\}$ is the set of types of individuals, $S_i (i \in I)$ is the finite set of strategies for each individual of type $i, M = (m_1, m_2, \ldots, m_n)$ specifies the number of individuals of each type who are matched in each matching situation, and π_i: $\times_{j \in I} S_j^{m_j(i)} \times S_i \to \Re$ where $m_i(i) = m_i - 1$ and $m_j(i) = m_j$ if $j \neq i$ is a payoff function for each individual of type i, where a typical value $\pi_i(s_1^1, \ldots, s_1^{m_1}, \ldots, s_i^1, \ldots, s_i^{m_i - 1}, \ldots, s_n^{m_n}; s_i)$ is the payoff for individual of type i when he/she takes s_i, while others take $(s_1^1, \ldots, s_n^{m_n})$. This somewhat awkward definition of the domain will simplify notations in the sequel. We assume that π_i is invariant with respect to permutation of strategies among the same type, i.e., among $s_j^1, \ldots, s_j^{m_j(i)}$. We bear in mind the interpretation according to which each $i \in I$ consists of a sufficiently large number of individuals who are anonymous and are matched randomly in each instance; without this interpretation, the definitions in the following sections will have little validity. Let $F_i \equiv \Delta(S_i)$ be the set of probability distributions over S_i, i.e.,

$$F_i \equiv \Delta(S_i) = \left\{ f_i \colon S_i \to \Re \mid \sum_{s_i \in S_i} f_i(s_i) = 1, \quad \text{and} \quad f_i(s_i) \geqslant 0 \text{ for all } s_i \in S_i \right\}.$$

We may call $F \equiv \times_{i \in I} \Delta(S_i)$ the class of strategy profiles and $f \equiv (f_1, \ldots, f_n) \in F$ a strategy profile. In considering the dynamic adjustment process, the current strategy profile will often be referred to as a behavior pattern. F is considered as $(\sum_{i \in I} |S_i| - n)$-dimensional space on which Euclidean norm, $\|\cdot\|$, and linear operations are defined. Given a strategy profile $f \in F$, the expected payoff for an individual of type i $(i \in I)$ if he/she takes a strategy $r_i \in S_i$ is:

$$\Pi_i(f; r_i) = \sum_{s \in \times_{j \in I} S_j^{m_j(i)}} \prod_{j \in I} \prod_{k=1}^{m_j(i)} f_i(s_i^k) \pi_i(s; r_i).$$

Let $Br_i(f)$ be the set of pure strategies for individuals of type $i \in I$ that are best responses to f, i.e.,

$$Br_i(f) = \operatorname*{argmax}_{r_i \in S_i} \prod_i (f; r_i).$$

Given $G \subset F$, we denote $Br_i(G) \equiv \bigcup_{g \in G} Br_i(g)$.

Let a function $[\cdot]$: $S_i \to \Delta(S_i)$ $(i \in I)$ satisfy $[s_i](s_i) = 1$ for all $s_i \in S_i$. The ε-neighborhood of a strategy profile f, denoted by $U_\varepsilon(f)$, is the set of strategy profiles g the distance of which from f in the Euclidean norm is less than ε.

3. SOCIAL STABILITY AND CYCLICALLY STABLE SETS

This section defines and discusses the concepts of social and cyclical stability. First of all, the definition of Nash equilibrium is given.

Definition: A strategy profile $f^* \in \times_{i \in I} \Delta(S_i)$ is a *Nash equilibrium* if f^* is a best response to f^* itself.

To capture the idea of social stability, we consider the following three points: (1) there are no strategic considerations such as retaliation; (2) unlike a deviation made by a single player, a change in behavior pattern is likely to be continuous; and (3) each player's ability to recognize the current situation is limited. To express these points, we introduce the notion of ε-accessibility.

DEFINITION: Given $\varepsilon > 0$ and strategy profiles f and g, g is *ε-accessible* from f if there exist a continuous function p: $[0,1] \to F$ differentiable from the right, a function h: $[0,1] \to F$ continuous from the right, and $\alpha \in [0,\infty)$ such that

$$p(0) = f, \quad p(1) = g,$$

and for each $t \in [0,1)$

$$(d^+/dt)p(t) = \alpha(h(t) - p(t)), \quad \text{and}$$

$$h(t) \in \times_{i \in I} \Delta\{Br_i[U_\varepsilon(p(t))]\}.$$

The definition says that in case of $\alpha > 0$, a behavior pattern moves in the direction of a convex combination of best responses to some strategy profiles which are in the ε-neighborhood of the behavior pattern, and it stays at the same place only if the behavior pattern is a best response to another one which is in the ε-neighborhood of itself. By including the case of $\alpha = 0$, we assure that a strategy profile is always ε-accessible from itself.

The interpretation of this definition is that only small and equal portions of individuals in each type realize the current behavior pattern and change their behavior pattern to another which is a best response to it. In doing so, there is a limitation on the ability of recognizing the current behavior pattern, so that its change may not be directed toward a best response to it; rather, it is only assumed that the direction is a best response to a possibly different behavior pattern which is in the ε-neighborhood of the current one. We may call the function p an ε-accessible path from f to g. Using this, accessibility from one strategy profile to another is defined.

DEFINITION: For two strategy profiles f and g, g is *accessible* from f if there exist sequences $\{\varepsilon_n\}_{n=1}^{\infty}$ in $(0, +\infty)$ and $\{g^n\}_{n=1}^{\infty}$ in F convergent to 0 and g respectively such that g^n is ε_n-accessible from f for all n.

Now, we are in a position to present the definition of cyclical stability.

DEFINITION: A nonempty subset F^* of $\times_{i \in I} \Delta(S_i)$ is *cyclically stable* if no $g \notin F^*$ is accessible from any $f \in F^*$, and every f^* in F^* is accessible from all f in F^*.

A strategy profile $f^* \in \times_{i \in I} \Delta(S_i)$ is called a *socially stable strategy* (*SSS*) if $\{f^*\}$ is cyclically stable.

A cyclically stable set (CSS) is stable in the sense that once the actual behavior pattern falls into it, another strategy profile may be realized if and only if it is within the CSS. The interpretation of this concept is as follows: For a long time, individuals have sought better strategies. After they search all the alternatives and acquire almost complete knowledge about the behavior pattern of other individuals, the actual behavior pattern may move within a CSS but never leave it. The term "cyclically stable" stems from the intuitive notion of cycles within the CSS. However, the paths may, of course, be much more complicated, especially when there are tie situations, in which case the behavior pattern may fluctuate arbitrarily along a continuum of strategy profiles.

Before we present the properties of CSS's, we present some important properties of the notion of accessibility, which are summarized in the following two lemmata.

LEMMA 1: *Suppose that $\{g^n\}_{n=1}^{\infty}$ is a sequence of strategy profiles all of which are accessible from $f \in F$. If g^n converges to $g \in F$, then g is accessible from f.*

PROOF: Let there be given $\{g^n\}_{n=1}^{\infty}$, f, and g as above. For each g^n, there exists a sequence (g^{nk}) such that g^{nk} is in the $1/k$-neighborhood of g^n and is $1/k$-accessible from f. Take the diagonal sequence $(\mu^k) = (g^{kk})$. Then (μ^k) converges to g and μ^k is $1/k$-accessible from f. Thus, g is accessible from f. *Q.E.D.*

LEMMA 2: *If $\tilde{g}$ is accessible from g which is accessible from f, then $\tilde{g}$ is accessible from f.*

PROOF: Suppose that $\tilde{g}$ is accessible from g and that g is accessible from f. Then there exists a sequence (g^n) converging to g such that g^n is $1/n$-accessible from f. Given $\delta > 0$, there exists $\hat{n}$ such that $g^n \in U_\delta(g)$ for all $n > \hat{n}$. Since $\tilde{g}$ is accessible from g, there exists a δ-accessible path from g to $\tilde{g}' \in U_\delta(\tilde{g})$, denoted by p. We construct a 2δ-accessible path from g^n to $\tilde{g}'' \in U_{2\delta}(\tilde{g})$, denoted by q, by using p. Since p is a δ-accessible path from g to $\tilde{g}'$, p is a solution to the problem:

$$(d^+/dt)p = \alpha_0(h^0 - p), \qquad p(0) = g,$$

for some $\alpha_0 \geqslant 0$ and a function h^0 continuous from the right on $[0,1]$. Since h^0 is continuous from the right, it has no more than a countable number of discontinuity points (see Gilboa and Matsui for a proof). Consider the problem: find a continuous q such that

$$(d^+/dt)q = \alpha_0(h^0 - q) \qquad \text{with} \quad q(0) = g^n.$$

By a well known theorem (see, e.g., Coddington and Levinson (1955, pp. 75–78)), such a q exists and is unique. Moreover, since h^0 is continuous from the right, $(d^+/dt)q$ equals $\alpha_0(h^0 - q)$ even at the discontinuity points of h^0.

Now, since $\|p(0) - q(0)\| < \delta$ holds, and p is a δ-accessible path, it is sufficient to show that $\|p(t) - q(t)\|$ is nonincreasing in t. If $\alpha_0 = 0$ the claim trivially holds, so suppose $\alpha_0 > 0$. First, we have

$$(d^+/dt)(p - q) = \alpha_0(h^0 - p) - \alpha_0(h^0 - q) = -\alpha_0(p - q).$$

Then we have

$$\begin{aligned}\|p(t+\tau) - q(t+\tau)\| &\leqslant \|\{p(t) - q(t)\} + (d^+/dt)(p(t) - q(t))\tau\| + o(\tau)\\ &= \|(1 - \alpha_0\tau)\{p(t) - q(t)\}\| + o(\tau),\end{aligned}$$

which is smaller than $\|p(t) - q(t)\|$ for sufficiently small $\tau > 0$. Thus, there exists $\tilde{g}'' \in U_{2\delta}(\tilde{g})$ which is η-accessible from f where $\eta = \max\{2\delta, 1/n\}$. This is true for all $n > \hat{n}$, and δ is arbitrary. Therefore, $\tilde{g}$ is accessible from f. *Q.E.D.*

4. PROPERTIES OF CYCLICALLY STABLE SETS

In this section, we prove that CSS's exist. Also, we will see the relationship between Nash equilibrium on the one hand and cyclically stable set and socially stable strategy on the other.

Existence

Before we state and prove the existence theorem for CSS, we denote by $R(f)$ the set of strategy profiles which are accessible from f, i.e., given $f \in \times_{i \in I} \Delta(S_i)$,

$$R(f) = \{g \in \times_{i \in I} \Delta(S_i) \mid g \text{ is accessible from } f\}.$$

In the proof, we make use of Zorn's lemma and the lemmata presented in the previous section.

THEOREM: *Every game has at least one cyclically stable set.*

PROOF: First, observe that $R(f)$ is nonempty for any $f \in F$; that, by Lemma 1, $R(f)$ is closed for any f; and that, by Lemma 2, $f' \in R(f)$ implies $R(f') \subset R(f)$.

Next, we consider the family of sets $\{R(f)\}_{f \in F}$ and define the inclusion $\subset$ as a partial order on them. Take any family $\{f^\alpha\}_{\alpha \in A}$ of strategy profiles such that for any α and β in A, either $R(f^\alpha) \subset R(f^\beta)$ or $R(f^\alpha) \supset R(f^\beta)$ holds. Consider $\bigcap_{\alpha \in A} R(f^\alpha)$, which is nonempty since the $R(\cdot)$'s are compact. Choose any f in $\bigcap_{\alpha \in A} R(f^\alpha)$ and recall that $R(f) \subset R(f^\alpha)$ holds for all $\alpha \in A$. Hence, $R(f)$ is a lower bound on the $R(f^\alpha)$'s. Therefore, by Zorn's lemma, there exists a minimal element $R^* = R(f^*)$ among the $R(\cdot)$'s. It is not empty because all the sets $R(f)$'s are nonempty.

We now claim that $R(f^*)$ is a CSS. Indeed, for any $f \in R(f^*)$ Lemma 2 implies that $R(f) \subset R(f^*)$. On the other hand, $R(f) \supset R^*$ holds for any f in R^* since R^* is a minimal element. Thus, $R(f) = R^*$ holds, which implies that every point in R^* is accessible from any point in R^*, and no point outside R^* is accessible from any point in R^*. *Q.E.D.*

By a similar argument, we can prove that for any strategy profile f, there exists a cyclically stable set any element of which is accessible from f. That is to say, the "domain of attraction" of all the cyclically stable sets is the whole space of mixed strategies (where a point f is said to be attracted to a CSS F^* if there *exists* $g \in F^*$ which is accessible from f; obviously, f may be attracted to several CSS's.) It is also worth noting that every CSS and its domain of attraction are closed and connected, that it is invariant with respect to sequential elimination of strictly dominated strategies and redundant strategies (for detail, see Gilboa and Matsui (1989)). We also note here that CSS's are neither upper nor lower hemi continuous with respect to the game payoffs.

Nash Equilibrium and Social Stability

We first have the following proposition.

PROPOSITION: *Any socially stable strategy is a Nash equilibrium.*

PROOF: Suppose that a strategy profile f is not a Nash equilibrium. Then there exist $\delta > 0$ and $\hat{s}_i \in S_i$ for some $i \in I$ such that any strategy profile in $U_\delta(f)$ takes $\hat{s}_i$ with probability of at least δ and $[\hat{s}_i] \notin Br_i(U_\delta(f))$ holds. Then for any $\varepsilon > 0$, there exists an ε-accessible path p which reaches the boundary of $U_\delta(f)$ since the speed of decrease in $p_i(t)(\hat{s}_i)$ is positive and bounded away from zero. Thus, there is a strategy profile in the boundary of $U_\delta(f)$ which is accessible from f since the boundary is sequentially compact. Hence, f cannot be a socially stable strategy. *Q.E.D.*

Next, we define a *strict Nash equilibrium* as a strategy profile f^* such that $Br(f^*) = \{f^*\}$, i.e., f^* is a profile of strategies which are strictly better responses to f^* than any other strategies. Then any strict Nash equilibrium is a socially stable strategy since for sufficiently small $\varepsilon > 0$, the set of the best response directions consists only of itself. Note that the converse is not true in general. In the game "matching pennies," for example, the mixed strategy Nash equilibrium is a socially stable strategy; on the other hand, it is not a strict Nash equilibrium (recall that any mixed strategy profile cannot be a strict Nash equilibrium).

Type 1

		L	C	R
Type 1	L	2, 2	1.2, 1.2	− 1, 3
	C	1.2, 1.2	1, 1	.2, .2
	R	3, − 1	.2, .2	0, 0

FIGURE 2

The concept of cyclically stable set is not directly related to that of Nash equilibrium. Though socially stable strategy is always a Nash equilibrium, each Nash equilibrium may be in some CSS or outside any of the CSS's. We proceed to show an example of a game which has no Nash equilibrium inside any CSS. Consider one-type game with two individuals matching in Figure 2. This game has a unique Nash equilibrium, $(\frac{1}{4}[L] + \frac{1}{2}[C] + \frac{1}{4}[R], \frac{1}{4}[L] + \frac{1}{2}[C] + \frac{1}{4}[R])$, if we regard it as a two-person game. In the following, we let (p, q, r) stand for $(p[L] + q[C] + r[R])$. We will find a CSS and then show that it is accessible from the unique Nash equilibrium, which does not belong to it. This will also prove that the Nash equilibrium does not belong to any other CSS. Figures 3 and 4 show the simplex of strategy profiles. In these figures, the vertex L of the triangle stands for the strategy profile $[L]$ and so on. The line segment AD indicates that if a strategy profile is on this line, then the pure strategies L and C yield the same expected payoff to the individuals. Similarly, on BE, individuals are indifferent between C and R, and on $C'F$—between L and R. Therefore, the area $ACBC'N$ is the one in which an individual prefers to take L; $C'LDEN$ is for R; and $ERFAN$ is for C. Finally, N is the Nash equilibrium.

In this game, a behavior pattern which differs from N swirls around it indefinitely without reaching any pure strategy profile. In fact, one may find that if the behavior pattern is inside PQR' of Figure 3 (or 4), then it follows an expanding cycle converting to PQR', and if it is outside PQR', it follows one shrinking to PQR', where $P = (.4, .5, .1)$, $Q = (.16, .2, .64)$, and $R' = (.04, .8, .16)$. If the behavior pattern is on PQR', then an ε-accessible path remains in some band around PQR' (see three triangular movements in Figure 4, in which dotted lines show ε-perturbation, that is, between dotted lines near P, for instance, both $(1, 0, 0)$ and $(0, 0, 1)$ are best response directions) and the bank shrinks to PQR' as ε tends to zero. Therefore, PQR' is a cyclically stable set. Since no matter

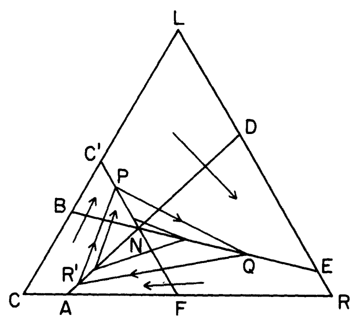

FIGURE 3

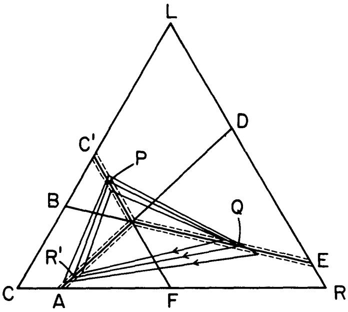

FIGURE 4

how small ε is, PQR' is accessible from N, there is no CSS which contains the Nash equilibrium in this game.

We do not view this phenomenon as a flaw of the concept of CSS; rather, it seems to us as criticizing the Nash equilibrium concept. To the extent that one finds the dynamic process presented above as reasonable, one is led to believe that Nash equilibrium may not be the appropriate tool for analysis of the evolution of economic behavior in large populations.

JL Kellogg Graduate School of Management, Northwestern University, Leverone Hall, Evanston, IL 60208, U.S.A.

Manuscript received April, 1989; final revision received April, 1990.

REFERENCES

AUMANN, R. J. (1987): "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1–18.

BINMORE, K. (1987): "Modeling Rational Players I," *Economics and Philosophy*, 3, 179–214.

——— (1988): "Modeling Rational Players II," *Economics and Philosophy*, 4, 9–55.

BRANDENBURGER, A., AND E. DEKEL (1987): "Rationalizability and Correlated Equilibria," *Econometrica*, 55, 1391–1402.

CODDINGTON, E. A., AND N. LEVINSON (1955): *Theory of Ordinary Differential Equations*. New York: McGraw-Hill.

GILBOA, I., AND A. MATSUI (1989): "Social Stability and Equilibrium," CMS-DMS No. 819, Northwestern University.

KALAI, E., AND D. SAMET (1984): "Persistent Equilibria in Strategic Games," *International Journal of Game Theory*, 13, 129–144.

KANEKO, M. (1987): "The Conventionally Stable Sets in Noncooperative Games with Limited Observations I: Definitions and Introductory Arguments," *Mathematical Social Sciences*, 13, 93–128.

KOHLBERG, E., AND J.-F. MERTENS (1986): "On the Strategic Stability of Equilibria," *Econometrica*, 54, 1003–1037.

MAYNARD SMITH, J., AND G. R. PRICE (1973): "The Logic of Animal Conflict," *Nature*, London, 246, 15–18.

MYERSON, R. B. (1978): "Refinement of the Nash Equilibrium Concept," *International Journal of Game Theory*, 7, 73–80.

NASH, J. (1951): "Non-cooperative Games," *Annals of Mathematics*, 54, 286–295.

SELTEN, R. (1975): "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4, 25–55.

SHAPLEY, L. S. (1964): "Some Topics in Two-Person Games," in *Advances in Game Theory*, ed. by M. Dresher, L. S. Shapley, and A. W. Tucker. Princeton: Princeton University Press, pp. 1–28.